

## Microspectrophotometric Analysis of Yellow Polyester Fiber Dye Loadings with Chemometric Techniques

Eric J. Reichard<sup>1</sup>, Edward G. Bartick<sup>2</sup>, Stephen L. Morgan<sup>3</sup> and John V. Goodpaster<sup>1\*</sup>

<sup>1</sup>Indiana University Purdue University Indianapolis (IUPUI), Department of Chemistry and Chemical Biology, Forensic and Investigative Sciences Program, 402 N. Blackford Street, LD 326, Indianapolis, IN 46202, reicharde@yahoo.com, [jvgoodpa@iupui.edu](mailto:jvgoodpa@iupui.edu)

<sup>2</sup>George Washington University, Department of Forensic Sciences, The George Washington University, 2100 Foxhall Road, NW, Washington, DC 20007, [ebartick@gwu.edu](mailto:ebartick@gwu.edu)

<sup>3</sup>University of South Carolina, Department of Chemistry and Biochemistry, Columbia, SC 29208, [morgansl@mailbox.sc.edu](mailto:morgansl@mailbox.sc.edu)

\*author to whom correspondence should be addressed

**Abstract.** Microspectrophotometry is a quick, accurate, and reproducible method to compare colored fibers for forensic purposes. Applying chemometric techniques to spectroscopic data can provide valuable information, especially when looking at a complex dataset. In this study, background subtracted and normalized visible spectra from ten yellow polyester exemplars dyed with different concentrations of the same dye ranging from 0.1-3.5% (w/w), were analyzed by agglomerative hierarchical clustering (AHC), principal component analysis (PCA), and discriminant analysis (DA). Systematic changes in the wavelength of maximum absorption, peak shape and signal-to-background ratio were noted as dye loading increased. In general, classifying the samples into ten groups (one for each exemplar) had poor accuracy (i.e., 51%). However, classification was much more accurate (i.e., 96%) using three classes of fibers that were identified by AHC as having low (0.10 – 0.20 wt%), medium (0.40 – 0.75 wt%), and high (1.5 – 3.5 wt%) dye loadings. An external validation with additional fibers and data generated by independent analysts confirmed these findings. Lastly, it was also possible to discriminating pairs of exemplars with small differences in dye loadings as a simulation of questioned (Q) versus known (K) comparisons.

**Keywords:** microspectrophotometry; forensic science; fibers; polyester; dye loading; multivariate statistics.

**Introduction.** The Locard Exchange Principle states that when two objects come into contact, there is always a transfer of material.<sup>1</sup> This principle is especially relevant to trace evidence involving textile fibers. Textile fibers can be identified and compared based on their macroscopic and microscopic characteristics, optical characteristics, chemical composition, and color.<sup>1,2</sup> In the comparison of transferred fibers during violent crimes, the fiber color is a very important point of comparison.

Color comparisons by eye are subjective and not always discriminating for fibers with similar dye colors. Spectral comparisons by UV/visible microspectrophotometry (MSP) are valuable in transferred fiber cases because the technique provides quick, non-destructive, and objective color comparisons for dyed fibers. Microspectrophotometry can discriminate between two colored fibers that are visually similar based upon the different chromophores in the dye's molecular structure.<sup>3-7</sup> Visual comparison of spectra is helpful, but can become difficult in comparisons of fibers that have the same dye composition, but different concentrations of individual dyes. Comparisons of this nature could be of importance for discriminating dyed fibers from the same manufacturer but different dye loadings or fibers that differ in the extent to which they have faded/photobleached. UV/visible spectra of textile fibers are dominated by the dyes present and tend to have broad peaks with few distinctive features to aid visual comparisons. Common analytical techniques employed for dye analysis include thin-layer chromatography (TLC),<sup>8,9</sup> liquid chromatography-mass spectrometry (LC-MS),<sup>10,11</sup> and capillary electrophoresis (CE).<sup>12</sup> All of these techniques are destructive to the sample, however. A detailed discussion of forensic textile fiber examination techniques is found in Robertson and Grieve<sup>13</sup> and Houck.<sup>14</sup>

Multivariate statistics uses measurements on multiple variables (*e.g.*, absorbances taken over a specified wavelength range) to identify patterns and groupings from large complex datasets more consistently than is possible by visual examination alone. We apply three techniques for analysis of colored fibers: agglomerative hierarchical clustering (AHC), principal component analysis (PCA), and discriminant analysis (DA). AHC groups samples based upon their relative similarity/dissimilarity, which is often expressed as a multivariate Euclidean distance.<sup>15</sup> An initial understanding of the natural groupings that

may be present in a dataset can be obtained from AHC. PCA is an unsupervised approach that reduces the dimensionality of the dataset by finding a smaller number of latent variables that explain maximum variance with linear combinations of the original variables (principal components, or PCs).<sup>15</sup> If most of the variation can be explained in the first few PCs, a plot of the data projected on the PCs may enable visualization of grouping relationships. On the other hand, discriminant analysis is a supervised approach. The known group membership of all spectra is employed to derive linear combinations of variables (canonical variates, CVs) that maximize the ratio of between-group variability to within-group variability. DA requires the data matrix to have more samples than features, which dimensionality reduction by PCA achieves.<sup>16</sup> DA generates a model that best discriminates the assigned groups in the original data, and which can predict the classification of new samples using the established model. Multivariate statistics have been applied to several relevant evidence types, including dyes,<sup>17</sup> inks,<sup>18</sup> automotive paint,<sup>19</sup> electrical tape,<sup>20</sup> and fire debris.<sup>21</sup>

The goal of our present work was to discriminate between different yellow polyester fiber visible spectra based solely upon their dye loadings using multivariate statistics. A dye loading is the concentration of a dye, usually in weight percent, applied to a fiber. Research has shown that visually similar yellow polyester fibers with different dye compositions can be discriminated based on PCA and DA of their UV/visible spectra.<sup>7</sup> However, no one has determined if fibers dyed with the same dye, but different dye loadings, can be reliably discriminated by their spectra alone using a chemometric approach. Being able to discriminate between dye loadings could provide a higher level of discriminating power to forensic fiber examiners in casework.

**Materials and Methods.** A large collection of textile fibers donated from textile companies to the University of South Carolina (Columbia, SC) was searched for a set of fibers with the same base polymer, fiber dye, cross section and thickness. This resulted in a sample set of ten yellow polyester exemplars dyed with various amounts of Dianix Yellow 5-6G (Disperse Yellow 114) dye. In general, polyester

fibers are among the most common synthetic fibers found in casework, however, target fiber studies have shown that coincidental matches to specific polyester fiber types are very rare.

Preliminary measurements using a Leica DM EP PLM (Leica Microsystems, Buffalo Grove, IL) and a Perkin Elmer Spectrum One FT-IR spectrometer with a universal ATR sampling accessory (Perkin Elmer, Waltham, MA) were performed to confirm that all of the polyester fibers had round cross-sections with similar diameters (20-27.5  $\mu\text{m}$ ). Cargille oils (R.P. Cargille Laboratories, Cedar Grove, NJ) were used to determine refractive indices by PLM. Michél-Levy charts were used to determine birefringence values. FT-IR spectra were acquired, based on averaging sixteen scans at a resolution of 4.00  $\text{cm}^{-1}$ , and recorded in percent transmittance over the wavenumber range of 4000-650  $\text{cm}^{-1}$ .

Table I shows the amount of dye applied, in weight percent, to each exemplar and the naming system employed for the study. Ten fibers from each exemplar were removed and mounted on glass microscope slides using Permout (Fisher Scientific, Fairlawn, NJ) mounting media. A QDI 2000 microspectrophotometer (CRAIC Technologies, San Dimas, CA) was used at a total magnification of 150 $\times$  in transmitted light mode. Calibration of the spectrometer with NIST traceable standards was performed before each use, along with Köhler illumination for the microscope. Autoset optimization, a dark scan, and a reference scan were employed prior to each sample scan. Fifty scans were taken at a resolution factor of five for each sample spectrum. Five spectra were taken at different locations along each fiber to account for variation in dye uptake. Hence, a total of 50 spectra were collected for each exemplar to provide sufficient information on dye loading variations within each exemplar. Visible MSP data was collected over the wavelength range of 350-800 nm.

Preprocessing techniques were employed before subjecting the data to statistical treatment. Background subtraction was performed on each spectrum by subtracting the minimum absorbance value for each sample from all absorbance values. Next, each background-subtracted spectrum was normalized to unit

vector length by dividing each absorbance value by the square root of the sum of squares of all absorbance values.

Statistical analyses were performed using XLSTAT Pro (AddinSoft, New York, NY), an add-in software for Microsoft Excel (Redmond, WA). For agglomerative hierarchical clustering analysis, the five replicate spectra for each fiber were averaged to produce a readable dendrogram. The proximity between two samples was measured by Euclidean distance, and Ward's Method was used for aggregation of samples.<sup>15</sup> This set of conditions has performed well in our prior work and it generates groups and dendrograms that were very consistent with the clustering seen in PCA.<sup>16-20</sup> The truncation line separating groups was set slightly higher than the most dissimilar exemplar's replicates. Dissimilarity values were determined by locating the node where all the replicates for each exemplar met. From that truncation line, exemplars were placed into classes.

PCA was conducted using the correlation matrix of the original variables (by mean centering the original variables and scaling to unit variance).<sup>15</sup> Factor loadings plots and scores plots were generated from the first two principal components. For subsequent discriminant analysis, three PCs were retained, based on visual examination of the scree plot of eigenvalues (data not shown).

An external validation was performed by separate analysts gathering data from three new fibers from each exemplar in the dataset. The same conditions were used as with the training dataset. Only PCA and DA were performed on these spectra from these fibers. PCA was performed to obtain factor scores of the validation set and subsequent DA was performed to predict to predict the classification of the new samples based upon dye loading.

Discriminant Analysis (DA) was performed with the assumption that the within-class covariance matrices of the classes were not equal (i.e., Quadratic Discriminant Analysis). In addition, the prior probabilities for class membership were considered to correct for any differences in the size of the classes.

Finally, PCA and DA were performed on pairs of dye classes from the original dataset. Dyed exemplars with dye loadings closest to each other were compared (e.g., A-B, B-C, C-D, D-E, *etc.*). In each case, the data from two dye classes were selected, followed by PCA and DA where only two classes are defined. In addition to classification accuracy, receiver operating characteristic (ROC) curves were generated as a part of the DA output to determine the performance of the model. ROC curves are generated by plotting the true positive and false positive rates associated with the model. The same pretreatments and conditions were used as with the training dataset.

**Results and Discussion.** Visual examination of the fibers revealed a slight difference in saturation between exemplars, however, discrimination of similar dye loadings was not possible. A preliminary examination of the exemplars by use of PLM and ATR-FTIR was conducted to confirm the identity of the fibers and to determine if the dye loading had any effect on the analysis. PLM was used to determine diameters, refractive indices, birefringence values, and signs of elongation. All exemplars had similar optical characteristics and dye loadings did not affect the results other than visual differences based on saturation. IR spectra were subjected to chemometric analysis and resulted in no significant differences between spectra of different fiber dye loadings. Library searches of the spectra were conducted, and they were highly correlated to exemplars of polyester.

Figures 1A, 1B and 1C show the averaged data for each exemplar in raw form (Figure 1A), after subtracting a constant background (Figure 1B) and after normalization (Figure 1C). Something that was immediately apparent in the spectra was a small but steady red shift in the wavelength of maximum absorption as dye loading increased (Figure 1A). This shift (from 422 nm to 435 nm) would seem to indicate that, at high concentrations, the dye itself is changing the chemical environment within the polyester. This shift is also independent of background or normalization and, hence, serves as a means of differentiation. Subtracting a constant background from all spectra reduces the effects of scattered light and brings the baseline for all spectra to zero absorbance (Figure 1B). This figure also illustrates a systematic change in the shape of the absorption – as reflected in the kurtosis and skewness of the peak.

In general, as dye loading increases, negative kurtosis increases (i.e., the absorption peak tails become thinner with respect to the central peak) and skewness decreases (i.e., the absorption peaks become more symmetric). These effects are independent of background or normalization and represent another potential means of discriminating spectra. Lastly, normalization accounts for differences in path lengths due to varying fiber diameters (Figure 1C). In this case, the spectra all pass through two “nodes” at 395 and 476 nm, but the differences in signal-to-background are very apparent. Fibers with a strong absorbance have a larger signal to background ratio than fibers with a weak absorbance. As can be seen in all spectra, the background spectrum decreases in absorbance from short to long wavelengths, which is due to increased light scattering at short wavelengths. The change in the standard deviation of the normalized absorption values is also presented in Figure 1C. The two distinct dips in this plot reflect the wavelengths at which the normalized absorbance for all spectra cross on either side of the peak maximum.

*Training set.* Spectra of the ten exemplars from Table I were subjected to AHC, PCA, and DA after background subtraction and normalization.

Agglomerative hierarchical clustering was performed first to detect outliers and to identify major classes of fibers. The intra-fiber spectra were averaged make it easier to visualize the classes in the dendrogram. This brought the number of replicates down from fifty to ten. The dendrogram in Figure 2 shows how AHC grouped the data, with increasing similarity to the left, and increasing dissimilarity to the right. A vertical truncation line was drawn to define the number of classes at a level of dissimilarity that was slightly higher than the highest level of dissimilarity between replicate spectra. Three distinct groups of spectra are suggested, representing low, medium, and high dye loadings. Class 1 are low dye loadings (0.1%, 0.2% w/w), and include exemplars A and B. Class 2 are medium dye loadings (0.4%, 0.5%, 0.75% w/w), and includes exemplars C, D, and E. Class 3 are high dye loadings (1.5%, 2.0%, 2.5%, 3.0%, 3.5% w/w), and includes exemplars F, G, H, I, and J. Based upon the within-class variance and the level of

dissimilarity at which the observations in each class are joined, Class 2 is the least homogenous while Class 1 is the most homogenous.

PCA was performed using all replicate spectra from each fiber exemplar. Projected into the space of the first two PCs, the variation of spectral data for individual exemplars over three classes accounted for 92.59% of the total variance of the data set. In general, the exemplars A-E are separated along the first PC, while exemplars F-J are separated along the second PC. Figure 3 shows the plot of the three classes produced from AHC. PCA and AHC show a consistent pattern suggesting that the three classes are due to three distinct levels of dye loading (low, medium, and high dye concentrations).

Whereas the vectors defining the principal components are eigenvectors of the data's variance-covariance matrix, the loadings plot (Figure 4) displays the contributions each wavelength makes to those eigenvectors, expressed as correlations between the original wavelengths and the resulting principal components. The first PC exhibits a strong negative correlation to wavelengths around the region of main absorption (408-453 nm) of the spectra. The second PC shows positive correlations around the leading and trailing edges of the fibers' spectral curve – the “nodes” as described above. In addition, a strong feature at 430 nm in PC2 correlates to the peak maximum. These regions of high correlation (positive and negative) define the relative contributions of these spectral regions in achieving class separation of the different fiber exemplars. Hence, the fibers with low dye loading were separated by overall absorption whereas the fibers with high dye loading were separated based on the shape of the spectra.

Discriminant analysis was performed in several ways on all the spectra. First, exemplars were assigned to their own classes and then to the three classes established by AHC. Leave-one-out cross validation was performed on the training set and a subsequent confusion matrix was produced. Here, DA is performed on all but one of the samples and a model is created. The sample that was left out is then classified based upon the model that was constructed. This step is repeated until every sample in the dataset is classified. However, cross validating the training set consisting of ten classes resulted in a low classification



accuracy of 50.80% (see Table II). Most misclassifications (i.e., 171 out of 254) occurred immediately off the diagonal of this matrix (e.g., between A and B, between B and C, between C and D, etc.). In addition, the classification accuracy generally decreased as dye loading increased. This decrease can be attributed to fibers with higher dye loadings producing spectra that are similar in shape but the transmittance approaches that of a dark scan.

Exemplars were also assigned to their AHC classes and DA was performed. Three classes were formed representing low, medium, and high dye loadings. The observation plot in Figure 5 had a total captured variance of 100% using two canonical variates. Although the 95% ellipses do overlap, the centroids for each class were not overlapped by the other ellipses. Leave-one-out cross validation produced an overall classification accuracy of 95.6% (Table III). The accuracy for each class did not vary significantly, ranging from 94% for Class 1 and 96.7% for Class 2.

*External Validation.* Three additional fibers from each exemplar were analyzed by PCA and DA to independently evaluate the accuracy of the classification model. Using DA, these additional samples were assigned to one of the ten exemplar classes for the first DA analysis, and then placed into one of the three AHC classes for the second DA analysis. Table IV shows the classification accuracies of the external validation. Classification accuracy when the fibers were placed into one of ten classes was mixed, ranging from 27% - 100%. Such a poor accuracy is not unexpected given that the cross-validation accuracy for the training set was also quite low. Overall, this result is a clear indication that classification of small differences in dye loading (e.g., 0.1 – 0.5%) is not reliable.

In contrast, classification accuracy when the fibers were placed into one of three classes determined by AHC ranged from 93% - 100%. Of the various external validation samples, only samples from Exemplars E and F were confused and this was limited to one sample misclassification each.

*Pair-Wise Comparisons.* PCA and DA were performed on several pairs of exemplars to determine whether the groups could be discriminated from each other. The results in Table V show classification

accuracy varies from 52% - 94% depending upon which two exemplars were being differentiated. The classification accuracy tended to decrease as the dye loadings of each pair increases. A similar trend was present in the area under curve for a ROC plot for each comparison, which varied from 0.644 for highly dyed fibers to 0.992 for lightly dyed fibers. While there are no official levels to gauge classification accuracy and AUC – one can use terms such as “excellent” ( $>90$ ), “good” ( $90 > x > 80$ ), “fair” ( $80 > x > 70$ ), “poor” ( $70 > x > 60$ ) and “fail” ( $< 60$ ). By these measures, six of the nine pair-wise comparisons are “good” or “excellent” in terms of classification accuracy and/or ROC AUC. Another important point is that pair-wise comparisons between exemplars that have the smallest difference in dye loading represent the most difficult scenario.

**Conclusions.** Chemometric treatment of UV/visible spectra from fibers with different dye loadings has shown to be a reliable and effective way of discriminating between yellow dye loadings when fibers were placed into classes of low, medium, and high dye loading. In addition, comparisons of two groups of fibers can provide discriminating information. Overall, for forensic fiber examiners, the chemometric approaches described in this work have the potential to provide an objective way of comparing UV/visible spectra of known and questioned fibers.

**Acknowledgements.** The authors acknowledge Dana Bors and Maria Diez for their contributions in acquiring spectra for the external validation of the training dataset. This project was supported by Award No. 2010-DN-BX-K220 from the National Institute of Justice, Office of Justice Programs, U. S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the Department of Justice.

## References

1. M.B. Eyring, B. D. Gaudette. An Introduction to the Forensic Aspects of Textile Fiber Examination, in: R. Saferstein (Ed.), Forensic Science Handbook, Vol. II, Prentice Hall, Englewood Cliffs, NJ, 2005, pp. 231-296.
2. S. Walkbridge-Jones. Microspectrophotometry for textile fiber color measurement, in: M. M. Houck (Ed.), Identification of textile fibers. CRC Press, Boca Raton, FL, 2009, pp. 165-180.

3. P. Adolf and J. Dunlop, Microspectrophotometry/colour measurement, in: J. Robertson, M. C. Grieve (Eds.), *Forensic Examination of Fibres*, second ed., Taylor and Francis, London, 1999, pp. 251-287.
4. M. Eng, P. Martin, C. Bhagwandin. The analysis of blue fibers and their forensic significance, *J. Forensic Sci.*, 54(4), 2009, 841-845.
5. M.C. Grieve, T. Biermann, M. Davignon, The occurrence and individuality of orange and green cotton fibres, *Sci. Justice*, 43(1), 2003, 5-22.
6. R. Macrae, R.J. Dudley, K.W. Smalldon, The characterization of dyestuffs on wool fibers with special reference to microspectrophotometry, *J. Forensic Sci.*, 24(1), 1979, 117-129.
7. S.L. Morgan, A.A. Nieuwland, C.R. Mubarak, J.E. Hendrix, E.M. Enlow, B.J. Vasser, E.G. Bartick, Forensic discrimination of dyed textile fibers using UV-VIS and fluorescence microspectrophotometry, *Proceedings of the European Fibres Group Annual Meeting*, 25 May 2004, Prague, Czechoslovakia.
8. K.G. Wiggins, S.R. Crabtree, B.M. March, The importance of thin layer chromatography in the analysis of reactive dyes released from wool fibers, *J. Forensic Sci.* 41(6), 1996, 1042-1045.
9. R. Resua, P.R. DeForest, H. Harris, The evaluation and selection of uncorrelated paired solvent systems for use in the comparison of textile dyes by thin-layer chromatography, *J. Forensic Sci.* 26(3), 1981, 515-534.
10. M. Huang, J. Yinon, M. Sigman. "Forensic identification of dyes extracted from textile fibers by liquid chromatography Mass spectrometry (LC-MS)". *J. Forensic Sci.*, 49(2), 2004, 238-249.
11. L.M. Petrick, T.A. Wilson, W.R. Fawcett, High-performance liquid chromatography-ultraviolet-visible spectroscopy-electrospray ionization mass spectrometry method for acrylic and polyester forensic fiber dye analysis". *J. Forensic Sci.*, 51(4), 2006, 771-779.
12. C.R. Dockery, A.R. Stefan, A.A. Nieuwland, S.N. Roberson, B.M. Baguley, J.E. Hendrix, S.L. Morgan, Automated extraction of direct, reactive, and vat dyes from cellulosic fibers for forensic analysis by capillary electrophoresis, *Anal. Bioanal. Chem.*, 394, 2009, 2095-2103.
13. J. Robertson, M. Grieve (Eds.), *Forensic Examination of Textile Fibers*, second ed., Taylor and Francis, London, 1999.
14. M. M. Houck (Ed.), *Identification of textile fibers*. CRC Press, Boca Raton, FL, 2009.
15. K.R. Beebe, R.J. Pell, M.B. Seasholtz, *Chemometrics: A Practical Guide*, John Wiley and Sons, New York, 1998.
16. S.L. Morgan, E.G. Bartick, Discrimination of forensic analytical chemical data using multivariate statistics, in: R.D. Blackledge (Ed.), *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*, John Wiley and Sons, New York, 2007, pp 333-374.
17. J.A. Barrett, J.A. Siegel, J.V. Goodpaster, "Forensic Discrimination of Dyed Hair Color: II. Multivariate Statistical Analy, *J. Forensic Sci.*, 56(1), 2011, 95-101.

18. A. Kher, M. Mulholland, E. Green, B. Reedy. Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis, *Vib. Spectrosc.*, 40, 2006, 270-277.
19. B.K. Kochanowshi and S.L. Morgan, Forensic discrimination of automotive paint samples using pyrolysis-gas chromatography-mass spectrometry with multivariate statistics". *J. Chromatogr. Sci.*, 38, 2000, 100-108.
20. J.V. Goodpaster, A.B. Sturdevant, K.L. Andrews, E.M. Briley, L. Brun-Conti, Identification and comparison of electrical tapes using instrumental and statistical techniques II. Organic composition of the tape backing and adhesive". *J. Forensic Sci.* 54(2), 2009, 328-338.
21. L.J. Marshall, J.W. Mcllroy, V.L. McGuffin, R.W. Smith, Association and discrimination of diesel fuels using chemometric procedures, *Anal. Bioanal. Chem.*, 394, 2009, 2049-2059.

**Table I.** Training dataset with respective naming designations and dye loadings in weight percent.

<b>Fiber ID</b>	<b>Naming System</b>	<b>% Dye Applied (w/w)</b>
674	A	0.10
675	B	0.20
676	C	0.40
677	D	0.50
678	E	0.75
679	F	1.50
680	G	2.00
681	H	2.50
682	I	3.00
683	J	3.50

**Table II.** “Leave-one-out” cross-validation confusion matrix for classifying samples by exemplar. The differences in dye loading between exemplars ranged from 0.1 – 0.5%.

<b>From\to</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>	<b>I</b>	<b>J</b>	<b>Total</b>	<b>%correct</b>
<b>A</b>	46	4	0	0	0	0	0	0	0	0	50	92.0%
<b>B</b>	5	40	5	0	0	0	0	0	0	0	50	80.0%
<b>C</b>	0	4	23	18	5	0	0	0	0	0	50	46.0%
<b>D</b>	0	0	18	18	14	0	0	0	0	0	50	36.0%
<b>E</b>	0	0	1	10	33	6	0	0	0	0	50	66.0%
<b>F</b>	0	0	0	0	7	30	8	5	0	0	50	60.0%
<b>G</b>	0	0	0	0	0	9	27	9	2	3	50	54.0%
<b>H</b>	0	0	0	0	0	7	10	11	5	17	50	22.0%
<b>I</b>	0	0	0	0	0	1	10	17	8	14	50	16.0%
<b>J</b>	0	0	0	0	0	2	7	15	8	18	50	36.0%
<b>Total</b>	51	48	47	46	59	55	62	57	23	52	500	<b>50.8%</b>

**Table III.** Leave-one-out cross-validation confusion matrix for prediction of class membership based on the three classes suggested by AHC. The difference in the mean dye loading of class 1 and class 2 was 0.40%. The difference in the mean dye loading of class 2 and class 3 was 1.95%.

<b>From\ to</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>Total</b>	<b>%correct</b>
<b>1</b>	94	6	0	100	94.0%
<b>2</b>	3	145	2	150	96.7%
<b>3</b>	0	11	239	250	95.6%
<b>Total</b>	97	162	241	500	<b>95.6%</b>

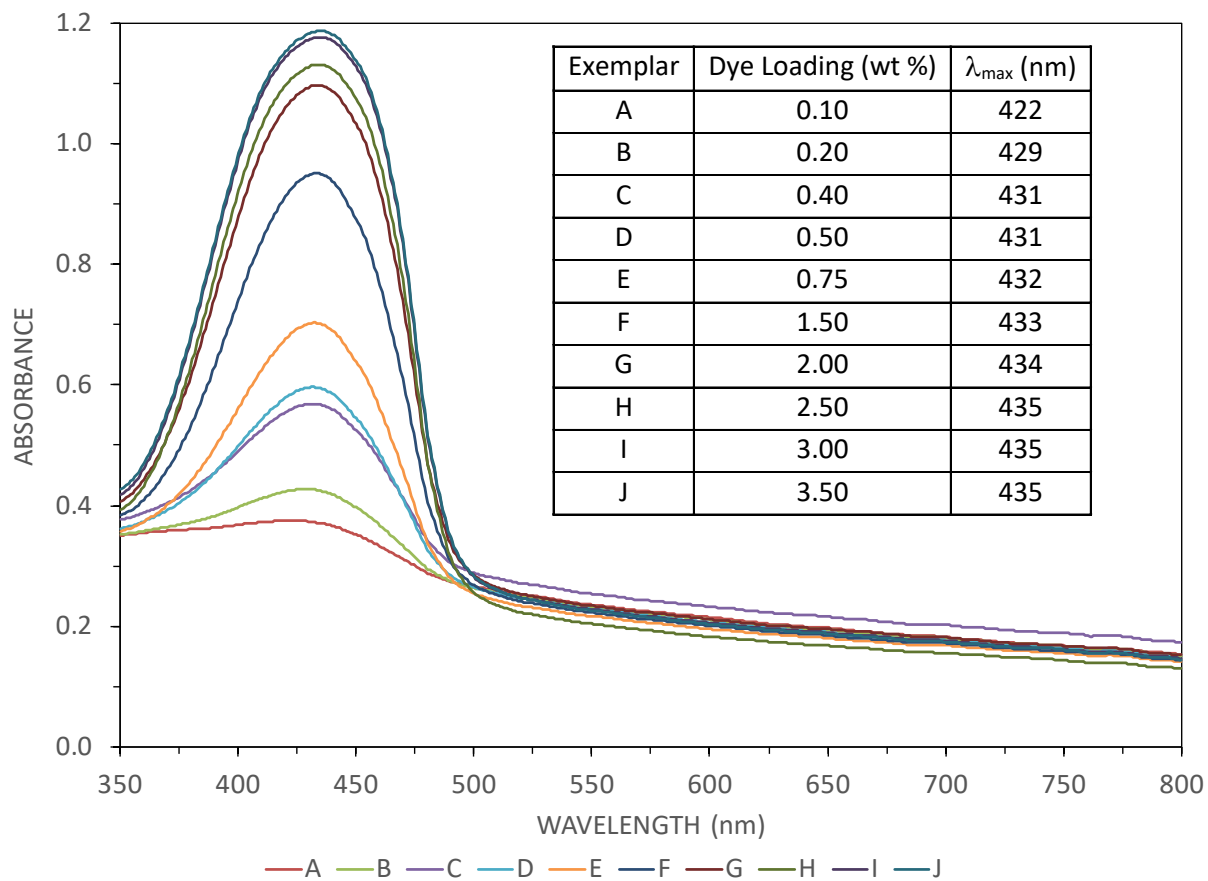
**Table IV.** External validation results: percentage of positive classifications for fibers grouped by exemplar and fibers grouped into respective classes generated from the AHC dendrogram.

Fibers Grouped by Exemplar			Fibers Grouped by AHC Class		
Additional Fiber	Class	Accuracy	Additional Fiber	Class	Accuracy
A	A	100 %	A	1	100 %
B	B	93.3 %	B	1	100 %
C	C	46.7 %	C	2	100 %
D	D	46.7 %	D	2	100 %
E	E	60 %	E	2	93.3 %
F	F	86.7 %	F	3	93.3 %
G	G	40 %	G	3	100 %
H	H	40 %	H	3	100 %
I	I	26.7 %	I	3	100 %
J	J	26.7 %	J	3	100 %

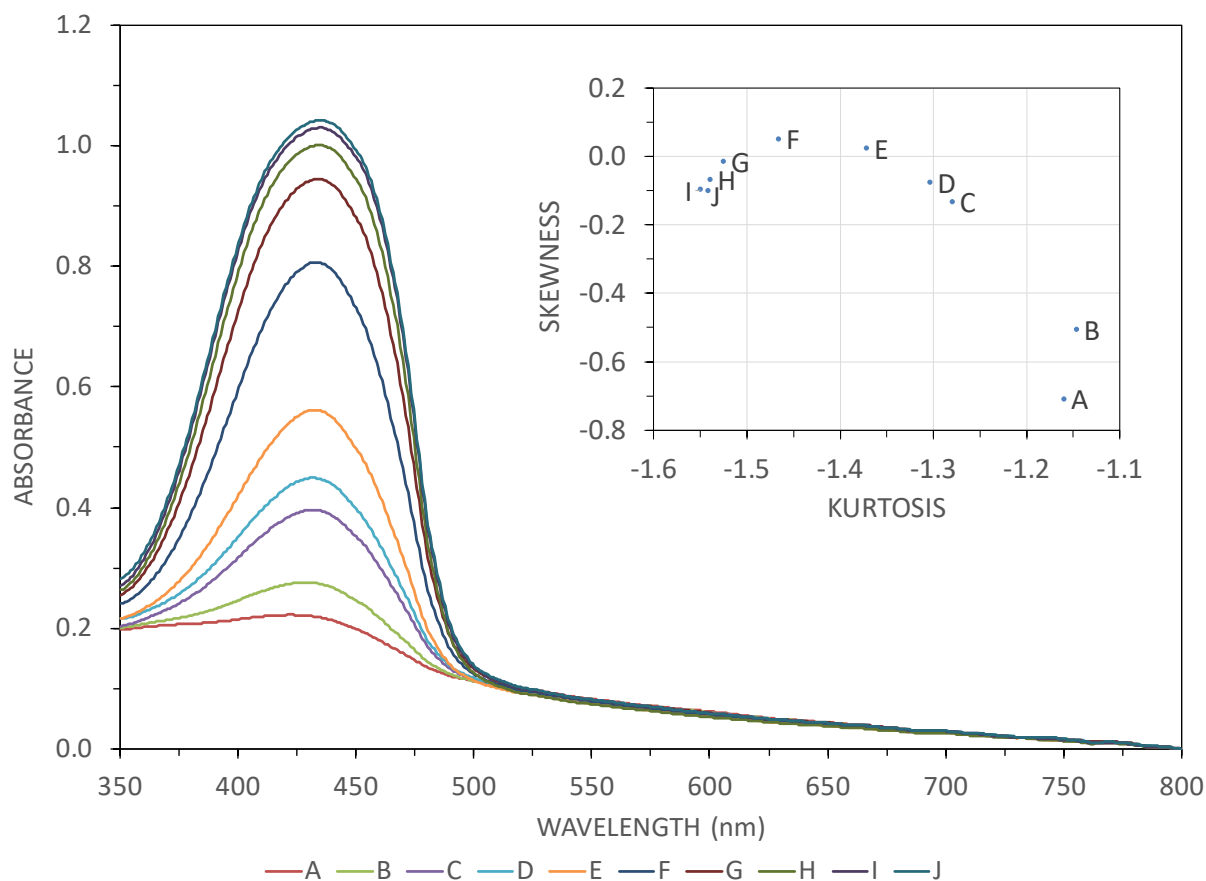


**Table V.** Classification accuracies and area under the curve (AUC) for ROC curves for the nine pair-wise comparisons. Comparisons shaded in grey do not have either a classification accuracy that exceeds 80% or an AUC that exceeds 0.8.

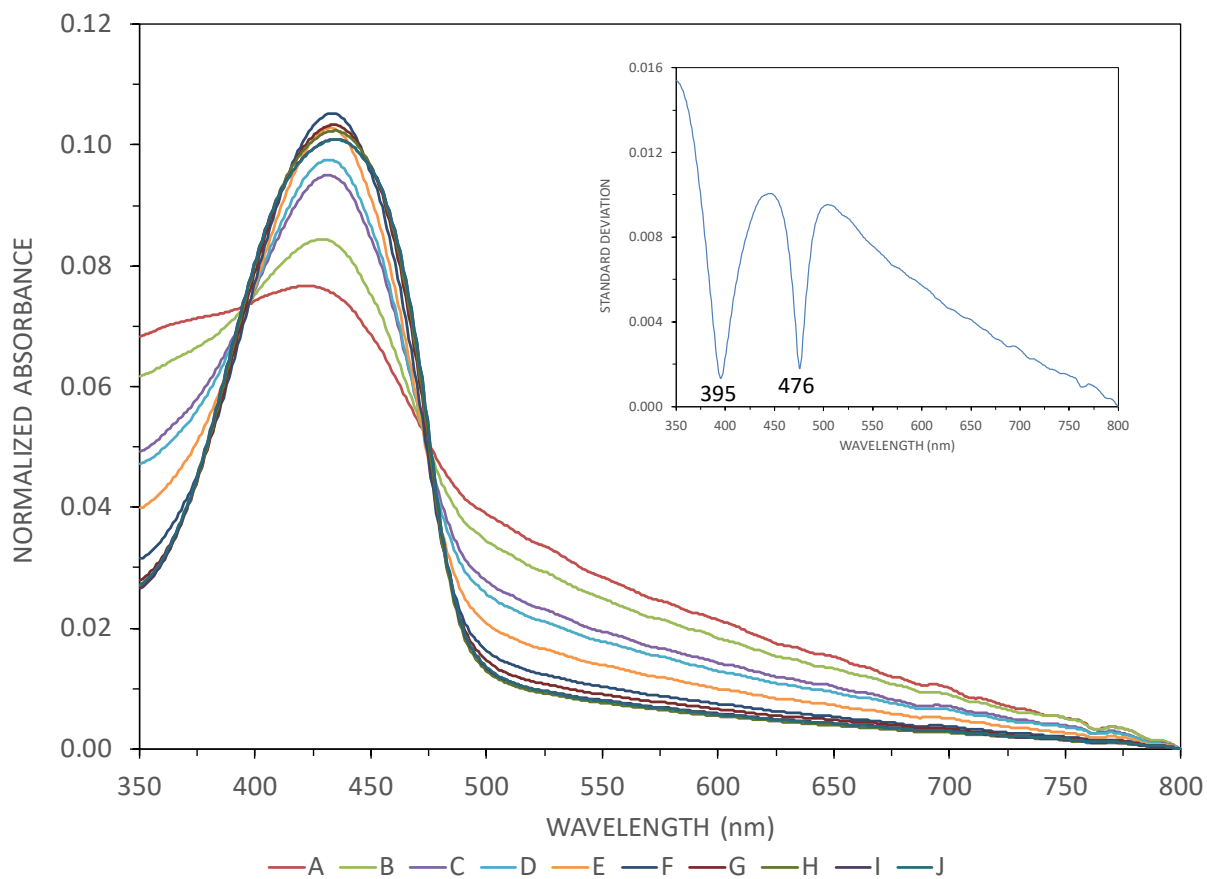
<b>Comparison</b>	<b>Classification Accuracy</b>	<b>ROC AUC</b>
<b>A-B</b>	92 %	0.992
<b>B-C</b>	94 %	0.992
<b>C-D</b>	64 %	0.720
<b>D-E</b>	82 %	0.936
<b>E-F</b>	89 %	0.982
<b>F-G</b>	83 %	0.918
<b>G-H</b>	75 %	0.845
<b>H-I</b>	57 %	0.721
<b>I-J</b>	52 %	0.644



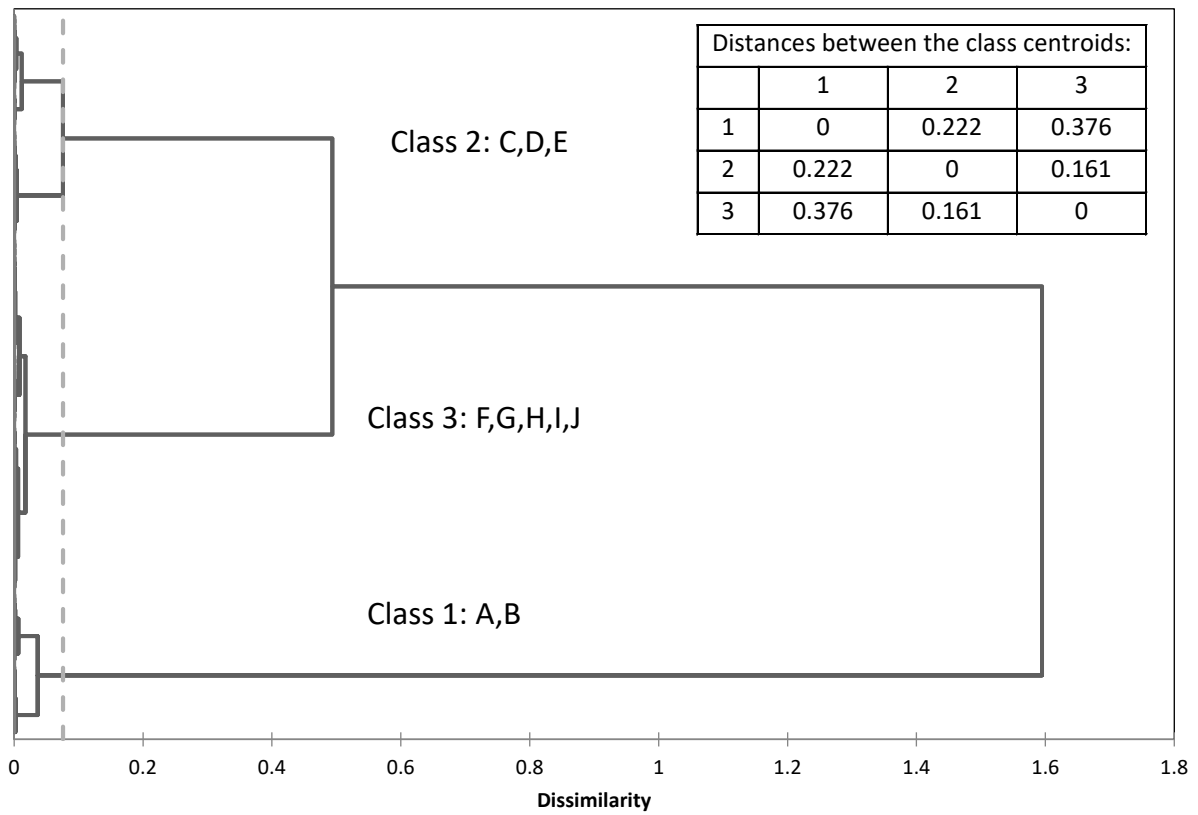
**Figure 1A.** Fiber exemplar spectra (raw data). The spectrum of each exemplar is an average of 10 fibers @ 5 scans per fiber (50 total scans)



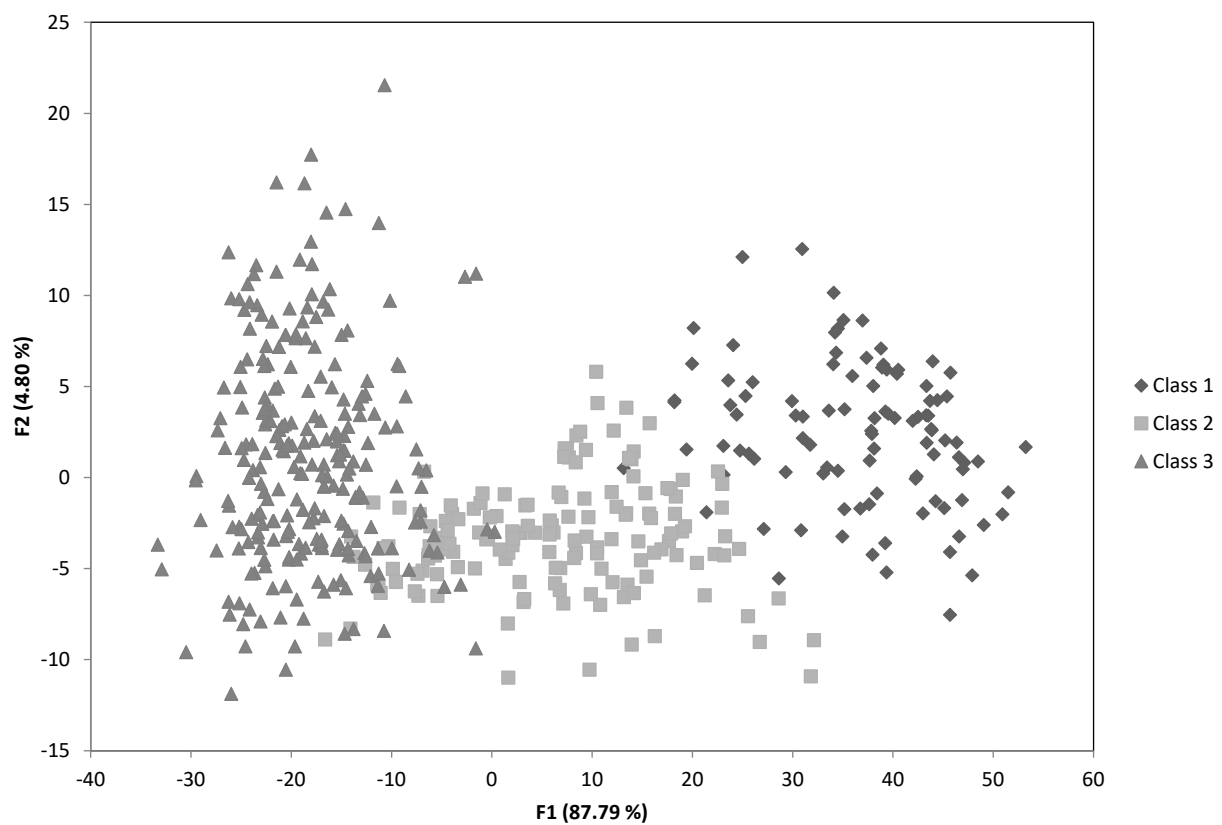
**Figure 1B.** Fiber exemplar spectra (background subtracted). The spectrum of each exemplar is an average of 10 fibers @ 5 scans per fiber (50 total scans). The inset figure illustrates the change in peak shape for each exemplar (calculated between 350 and 525 nm).



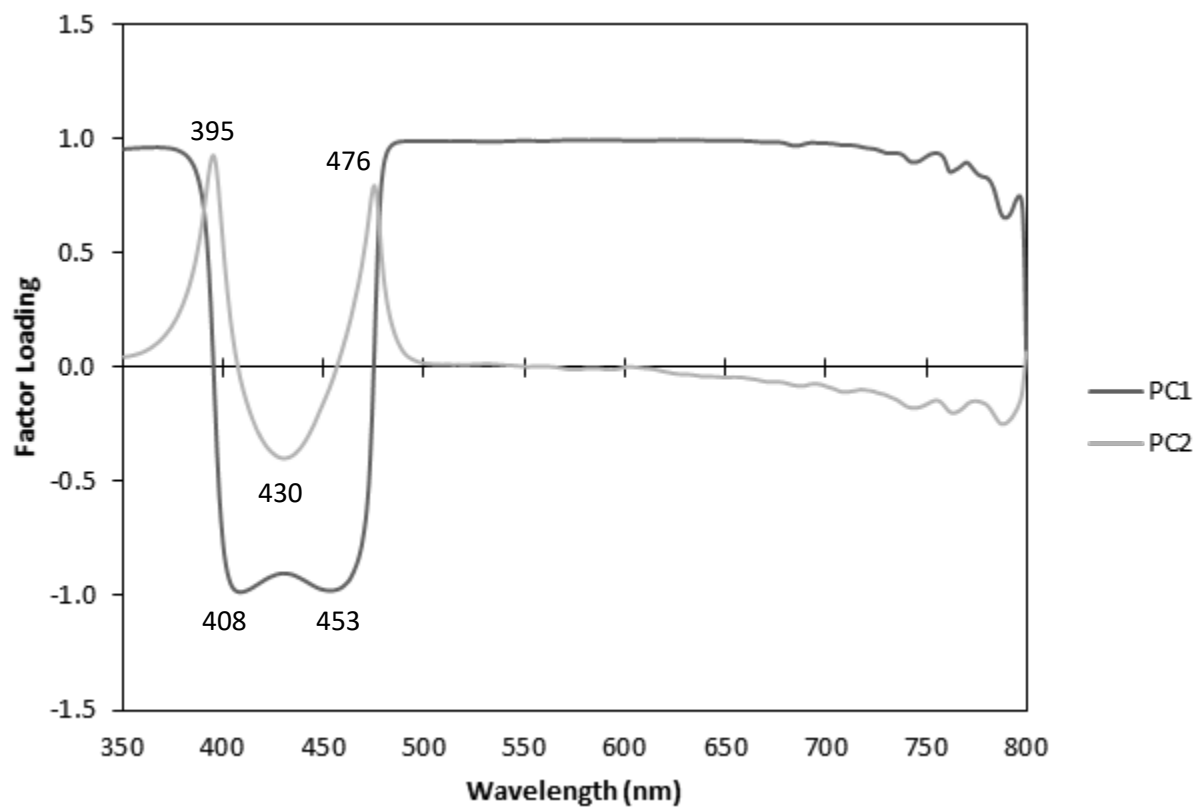
**Figure 1C.** Fiber exemplar spectra (background subtracted and normalized). The spectrum of each exemplar is an average of 10 fibers @ 5 scans per fiber (50 total scans). The inset figure depicts the standard deviation of the variables as a function of wavelength.



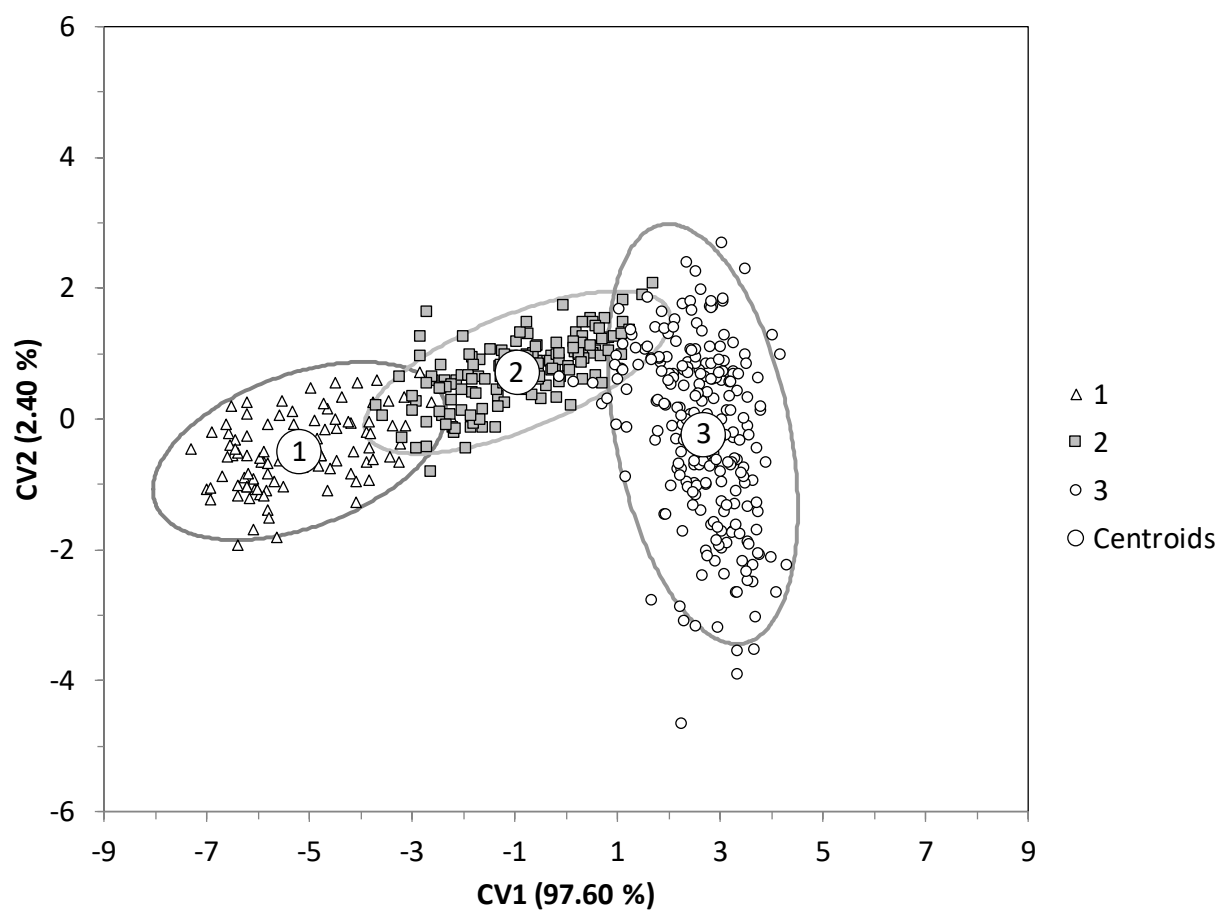
**Figure 2.** AHC dendrogram of the 10 exemplars from the training set (10 fibers per exemplar).



**Figure 3.** Projections of the three classes generated from the dendrogram (Class 1: A, B; Class 2: C, D, E; Class 3: F, G, H, I, J) onto the principal components shown in Figure 4.



**Figure 4.** Factor loadings plot for first two principal components. Large negative projections on PC1 are associated with large absorbance between 408 nm and 453 nm. Projections on PC2 are associated with the absorbance at the “nodes” of the normalized spectra (see Figure 1C) in relation to absorbance at the peak maximum (430 nm).



**Figure 5.** Scores plot of the first two canonical variates of three classes generated from the AHC dendrogram (Class 1: A, B; Class 2: C, D, E; Class 3: F, G, H, I, J).